Machine Learning
Exercise Sheet 3

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond
HiWi's: Khouloud Sallami, Maurício Camargo, Shayan Jawed, Sami Diaf
Information Systems and Machine Learning Lab
University of Hildesheim

Exercise 5: Logistic Regression (10 Points)

a) Describe the main difference between classification and regression problems!

b) A logistic regression is a binary classifier, where:

$$\hat{y}(x; \beta) = \begin{cases} 1 & \text{if} \quad \sigma(\beta^\top x) > 0.5 \\ 0 & \text{if} \quad \sigma(\beta^\top x) \le 0.5 \end{cases} \tag{1}$$

for $\sigma(t)$ being the sigmoid function:

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{2}$$

For the data:

x_1	x_2	y
2	4	1
3	3	1
-4	-2	0
-2	-6	0

perform **one iteration of gradient ascent**! Initialize all your parameters to zero and use a step size $\alpha = 0.5$!

c) Perform **one iteration of Newton**! Again, initialize all your parameters to zero and use a step size $\alpha = 0.5$!

Exercise 6: Multiclass Classification (10 Points)

a) Describe the differences between the 1-vs-rest and 1-vs-last approaches for multiclass classification. What are the positive, what are the negative aspects of each approach?

b) For the following data containg three labels:

x_1	x_2	y
4	2	1
4	4	1
4	-2	2
2	-4	2
-8	6	3
-2	4	3

three binary 1-vs-rest classifiers have been learned. The resulting model parameters are

$$\beta_1 = \begin{pmatrix} 1 & 1 & 2 \end{pmatrix} \qquad \beta_2 = \begin{pmatrix} 0 & 2 & -3 \end{pmatrix} \qquad \beta_3 = \begin{pmatrix} 1 & -4 & 1 \end{pmatrix} \tag{3}$$

Compute the predictions for all the data points and assign labels accordingly. What is the final misclassification rate?